2 Basic Approach (theory).

$\underbrace{\text{M-estimators}}_{\hat{\theta}}$ are solutions of the vector equation (iid case)

$$\sum_{i=1}^{n} \boldsymbol{\psi}(\mathbf{Y}_i, \boldsymbol{\theta}) = \mathbf{0}.$$

but what are they estimating? Some true parameter $\underline{\theta}_0$, where.

$$(*) \quad E_F\left[\underline{\Psi}(Y_1; \underline{\theta}_0)\right] = \int \underline{\Psi}(y; \underline{\theta}_0)\, dF(y) = 0 \quad \text{where} \quad Y_1 \sim F$$

Example (Sample Mean, cont'd): Recall we said $\theta = \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ is an M-estimator for $\psi(Y_i, \theta) = Y_i - \theta$. What is the true parameter?

The true parameter solves $\int (y - \theta)\, dF(y) = 0$

$$\Rightarrow \underbrace{\int y\, dF(y)}_{\text{population mean}} = \theta$$

Recall the 5-dimensional motivating example.

We said the α which maximizes the pairwise loglikelihood seems like it would be a good estimator for α_0.

We didn't show this.

To do this, we would need to use $(*)$

To arrive at the sandwich estimator, assume $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n \overset{iid}{\sim} \boldsymbol{F}$ and define

$$\boldsymbol{G}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\psi}(\boldsymbol{Y}_i; \boldsymbol{\theta}).$$

↑depend on n.

In the likelihood case:

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} \underbrace{\underline{\Psi}(\underline{y}_i;\underline{\theta})}_{\text{score contribution, deriv of log likelihood contribution}}}_{\text{mean derivative of log likelihood contributions.}}$$

Taylor expansion of $\boldsymbol{G}_n(\boldsymbol{\theta})$ around $\boldsymbol{\theta}_0$ evaluated at $\hat{\boldsymbol{\theta}}$ yields

$$0 = \underline{G}_n(\hat{\underline{\theta}}) = \underline{G}_n(\theta_0) + \underline{G}_n'(\theta_0)\underbrace{(\hat{\underline{\theta}} - \underline{\theta}_0)}_{b\times 1} + \underline{R}_n$$

definition of $\hat{\theta}$

bx1

bxb Jacobian

higher order terms "residual"

Rearranging :

$$-\underline{G}_n'(\underline{\theta}_0)(\hat{\underline{\theta}} - \underline{\theta}_0) = \underline{G}_n(\underline{\theta}_0) + \underline{R}_n$$

$$\hat{\underline{\theta}} - \underline{\theta}_0 = \{-\underline{G}_n'(\underline{\theta}_0)\}^{-1}\underline{G}_n(\underline{\theta}_0) + \underbrace{\{-\underline{G}_n'(\underline{\theta}_0)\}^{-1}\underline{R}_n}_{R_n^*}$$

$$\sqrt{n}\,(\hat{\underline{\theta}} - \underline{\theta}_0) = \underbrace{\{-\underline{G}_n'(\underline{\theta}_0)\}^{-1}}_{①}\underbrace{\sqrt{n}\,\underline{G}_n(\underline{\theta}_0)}_{②} + \underbrace{\sqrt{n}\,\underline{R}_n^*}_{③}$$

Define $\boldsymbol{A}(\boldsymbol{\theta}_0) = \mathrm{E}_F[-\boldsymbol{\psi}'(\boldsymbol{Y}_1, \boldsymbol{\theta}_0)]$.

2.1 Estimators for $\boldsymbol{A}, \boldsymbol{B}$

If the data truly come from the assumed parametric family $f(y; \boldsymbol{\theta})$,

One of the key contributions of M-estimation theory is to point out what happens when the assumed parametric family is not correct.

We can use empirical estimators of $\boldsymbol{A}$ and $\boldsymbol{B}$:

Example (Coefficient of Variation): Let $Y_1, \ldots, Y_n$ be idd from some distribution with finite fourth moment. The coefficient of variation is defined at $\hat{\theta}_3 = s_n/\overline{Y}$.

Define a three dimensional $\boldsymbol{\psi}$ so that $\hat{\theta}_3$ is defined by summing the third component. What is the vector valued function $\boldsymbol{\psi}$ which yields an M-estimator for the coefficient of variation?

What parameter vector is being estimated by the M-estimator?

What are the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$?

Write out the asymptotic variance, $\boldsymbol{V}$.

Assume Y_i are iid from a normal distribution with mean 10 and standard deviation 1. Calculate $V_{3,3}$. Assume you have a same of size 25 and you get an estimated coefficient of variation of 0.11. Give the asymptotic 95% confidence interval.